

# Improving Access to NBN data and products

Results summary (September 2015) – correct as of 0900, 10/09/2015

*By Rachel Stroud*  
*NBN Data Liaison Officer*

## **Executive summary**

1. As a first step in implementing the NBN Strategy 2015-2020 the NBN Secretariat sought views from all NBN Data Partners about a range of issues. These issues included:

- Data sharing globally via GBIF
- Non-native and invasive species
- Pollinators
- Data shared via an atlas platform
- Data licensing
- Data flow (especially via iRecord)

2. Main questions raised were:

- Definitions of 'non-native', 'pollinator'
- LERC business models
- More information about data licenses and implications
- Monitoring commercial use
- Could consultants be 'data partners'?
- Data resolution for GBIF and Atlas and for non-native, pollinator and historic

3. In general, Data Partners are keen to make changes with regards to opening up data on non-natives (92% Yes or Maybe), pollinators (77% Yes or Maybe), historic data (84% Yes or Maybe).

4. In general, Data Partners are keen to increase data flow between NBN members (94% Yes or Maybe)

5. In general, Data Partners are keen to share data via an Atlas (94% Yes or Maybe) and with GBIF (89% Yes or Maybe).

6. These results can be viewed the other way in that adding No and Maybe also achieves high scores. This provides a good launching point for opening up conversations with NBN Data Partners about Data sharing. Many of the comments and issues are either easily resolved or not relevant to this exercise. We will focus future workshops and discussions around resolving the remaining issues.

7. NBN has met with Ordnance Survey and has prepared a case for a Third Sector OS license. This has been provided to OS Head of Licensing (John Carpenter) and rejected by Ordnance Survey.

### **Total Responses = 67**

- 63 with data on NBN Gateway
- 4 without data on NBN Gateway

### **Q1. Increasing access to non-native data**

**Will your organisation make all non-native data open on the NBN Gateway?**

- Yes 59% (n=37)
- No 8% (n=5)
- Maybe 33% (n=21)

#### **Comments**

- Subject to recorder or owner's consent
- As long as any sensitivity issues are taken into account
- Do not have sufficient resources to upload these data to the NBN Gateway
- Access to high resolution data would be limited by DPA agreements with landowners
- Would make visible and accessible but not full resolution
- These are key data for commercial consultancies and funded research. Will provide this if a solution to funding LERCs is developed

#### **Questions to consider**

- What is definition of non native species
- Can we make it easier for users to separate these data (i.e a tick box on metadata)
- How can avoid duplication of datasets on NBN Gateway - is this an issue?
- What resolution would this be at?

### **Q2. Increasing access to pollinator data**

**Will your organisation make all data for pollinators open on the NBN Gateway**

- Yes 40% (n=25)
- No 12% (n=8)
- Maybe 37% (n=23)

#### **Comments**

- Subject to recorder and owner's consent
- As long as sensitive species taken into account
- Wish to retain control over access to data such that it does not undermine our current business model.

- Will make data available to those who request access to it - do not need to make open
- No funds available to prepare data for upload to NBN Gateway
- Would make visible and accessible but not at full resolution
- Would have to be able to flag unverified records
- Would not be appropriate for us to agree to make all data open without option for exception.

### Questions

- What is definition of pollinators
- Which species would these be
- Need to make simple to extract these data from current datasets
- LERCs should make this an option when uploading data to their systems

### Q3. Increasing access to historic data

Would you be willing to make all data that is older than 15 years open on the NBN Gateway?

- Yes 46% (n=29)
- No 16% (n=10)
- Maybe 38% (n=24)

### Comments

- Subject to recorder's and owner's permission
- Only for verified data
- Would prefer a long period of time (25/35 years)
- Some species should not be included which are sensitive and vulnerable for much longer than 15 years (e.g. badger setts/ many Dormouse sites have been recording for over 15 years and so releasing the old data fully would compromise the newer data also)
- Staff and resources are required to contact owners and gain consent
- Releasing such data would undermine current business model of LERCs. Our partnership and operating model does not allow for historic data to be treated any differently to our other holdings unless there were changes to our funding
- Want want to retain restriction on commercial
- Would not be appropriate for us to agree to make all data open without option for exception.
- Unlikely that recorder name would be visible as standard.

### Questions

- Need to know more details of the proposal
- Is this at full resolution?
- Interest in getting data from other sources eg in literature online
- Depends what is meant by "data". Occurrence data, probably yes. But not necessarily all attributes, and not necessarily all at capture resolution.

#### **Q4. Mobilising historic datasets**

**Do you have data sets, in digital and non-digital formats that you want to share with NBN nationally via the NBN Gateway, or regionally via Local Environmental Record Centres, but have not done so because of resourcing issues?**

- Yes 51% (n=32)
- No 33% (n=21)
- Maybe 14% (n=9)

#### **Comments**

- Would prefer that LERCs are funded to mobilise data. Needs supervision as this is what we are skilled to do.
- Money might be better spent on up-to-date field surveys rather than mobilising historical data
- There still seems to be little consideration given to the functions of LERCs outside of data handlers.
- Unfortunately, a remarkable number of these species sightings also lack at least one of the essential requirements of species, location, date & recorder, so are unsuitable for going on the NBN at a meaningful resolution!

#### **Why not been able to share**

- Lack of resource and money be it for data collection or data transcription and entry into useable electronic forms.
- Some datasets require significant data management before they could be uploaded

#### **How can we help**

- Additional resources (e.g. money to pay contractors) could help us mobilise the data more quickly.
- Considering sponsorship from organisations such as research institutions for mobilising data that would be useful for research purposes, in return for making the data freely available to those institutions.
- By identifying and addressing the issues which (a) would encourage data owners to share or waive rights to their data, and (b) LERC business models.

#### **Q5.Data access request response time**

**Are there reasons why the NBN should not show response time?**

- Yes 24% (n=15)
- No 73% (n=46)

#### **Comments**

- As long as clear that not all data are on Gateway. Normally refuse requests as only proportion of our total data holdings are on NBN Gateway so ask for people to come direct

- As long as local responses are also factored in to the statistics (a request may be turned down via the Gateway in order to provide the user with a more detailed response locally)
- Would be happy to have our inhouse target response time advertised (7 days/5days/same day)
- Sometimes NBN Gateway crashes and cannot respond to requests
- Receive a lot of inappropriate requests for data outside area, for sensitive spp. are sent as people don't read the data access conditions nor do they understand the filtering system.
- Consideration should be given to small organisations/individuals who may not have the time/resources to put into data management that larger organisations do.
- Follow up communications for more information or clarification often are ignored
- Would be useful to have an automated response from data providers who are unlikely to respond/holding acknowledgement

### Question

- Would this be the (mean) average response time or some other measure?

## Q6. Increasing data access between NBN members

If data management standards were met, would you be willing for all other NBN Data Partners to see, access and make use of your data in their work?

- Yes 51% (n=32)
- No 6% (n=4)
- Maybe 43% (n=27)

### Comments

- Subject to data owner's permission
- Need to ensure commercial/research users don't use this as way to circumvent paying for data searches
- If via Internal information request process
- If partners have signed data partner/sharing agreement
- As long as data owner acknowledged in any derived material
- Organisations shouldn't really get special privileges just because they are dataset administrators, because many other individuals and organisations may have contributed to the dataset they administer.
- Access to high resolution data limited by DPA agreements with landowners. This allows statutory agencies in England and Wales access only
- Unless data is for a specific project
- Unless restrictions are placed on the data
- As long as have information on the use data have been put to, irrespective of whether the user is an NBN partner or not.

- Would want to see the list of members and have opportunity to confirm that proposed use and access is appropriate – it may not be appropriate for all to have full access to sensitive species records.
- Would want assurances that the data would be for the member's own (non-commercial) use only: they should not pass it on to others, including building a database to provide data services that compete with the LERC and should not be able to make any charge to others for any product that has used the data in its creation

### Questions

- What is a data partner?
- Clarity needed here as it is not all about management standards, it can be about who, and why wants access.

### General comments

- A separate function, similar to the access requests, is needed to handle permission requests and keep a log of permissions granted and refused.
- Asking for permission is time consuming - if NBN could help here that would be great
- It would be good if there could be an automated alert to users of datasets if there is an amendment or update.
- One area that might benefit from clear standards is around best practice over the handling of sensitive records. A clearer set of guidelines to enable us all to grant access to sensitive records in an appropriate way could be useful across the board?
- Need to improve data flows generally
- Looking forward, there is a need to develop effective technical means to enable information users to have simple and fast access to non-sensitive data AND to levy charges so that those organisations which presently depend on charging for access to/use of data to fund (a) the validation and verification processes on which data quality depends, (b) support of local recording activities and (c) giving local context to data can continue to operate. Again there is a need for NBNT to help foster change.

## Q7.Creative Commons licenses within the NBN

**Do you administer datasets on the NBN Gateway that you would be willing to change to any one of these licenses?**

### Open Government License:

Yes	14%	(n=9)
No	27%	(n=17)
Maybe	3%	(n=2)

**Creative Commons with Attribution (CC-BY):**

Yes	19%	(n=12)
No	25%	(n=16)
Maybe	37%	(n=23)

**Creative Commons with Attribution, non-commercial (CC-BY-NC):**

Yes	38%	(n=24)
No	21%	(n=13)
Maybe	29%	(n=18)

**Comments**

- Consultation will be required with others (local groups, data owners, members etc)
- Sensitive species protection will need to be taken into account
- Will wait until a higher proportion of our data on NBN Gateway
- More clarification needed on exactly what these licenses allow, how commercial use will be controlled and to understand the implications
- These licenses are more suitable to organisations that produce data as a 'byproduct' of their remit, such as NE/EA. As a data service provider, we don't fund the upkeep of core datasets to then distribute to third parties to create a new service that we could be providing.

**General comments**

- There is a need to begin signposting the next step, where commercial users are able to access data easily for a fee which then ends up with the relevant provider and create the platform to make this workable, whilst increasing the speed and efficiency of commercial searches.
- From what I've read and heard, CC0 with T&Cs requiring attribution would be a more robust choice than CC-BY. I'm not sure how this would be monitored and enforced, neither am I sure how you would go about acknowledging the source of data (dataset administrator, recorder, determiner etc?). A DOI to a cached bundle of data as provided by GBIF would be ideal for researchers wishing to replicate methods and presumably this will be part of the new ALA model adopted in the UK.

**Questions**

- Would it be possible to mark all of our data at CC-BY-NC with the potential to grant a CC-BY commercial license on a case-by-case basis?
- More clarity would be useful around the term "commercial". I couldn't see anywhere a clear definition of the term "commercial".

- We'd still like some clarity on audit of data download from NBN. Not sure that any of these licenses say anything about informing us of the access being made.

## **Q8. Atlas infrastructure for biodiversity data in the UK**

**In principle would you be willing for you data to be held on a UK Atlas?**

Yes	65%	(n=41)
No	5%	(n=3)
Maybe	29%	(n=18)

### **Comments**

- Depends on the resolution the data are available at
- With the proviso above - blurred data only.
- Consultation with local groups required
- Access to high resolution data limited by DPA agreements with landowners.
- Access to data through any such system needs to respect the business models of the contributing data partners.

### **Concerns**

- Concerned that the web services, in which we have made significant investment and are regular users, will be lost. If this was the result we would have to stop using the Gateway.
- Not under that creative commons licence or we would have to restrict data we hold on the NBN to only species such the very common species
- As long as the core data providers such as LRCs and recording schemes are recognised for their importance in mobilising data and are thus funded in a long term sustainable way to continue improving and maximising the data flow
- I don't think the Atlas should have NC licenses because it is impossible to enforce this (not particularly easy to define it either!). The current ALA model does not have access controls. You can't enforce NC licenses without access controls (users will just say 'oh, i didn't realise that type of use would be defined as commercial'). Or are there plans to develop access controls for the UK version of ALA?
- Would want to be clearly highlighted as the data provider and that the organisation must be contacted for up to date information (holdings would only be updated e.g. annually) – or for commercial needs.

### **Questions**

- How does this differ from the how data is displayed/used by the NBN now?



- Will the new Atlas replace or run alongside the 'old' Gateway.

### Q9. Data flow from iRecord

When biological data, collected by volunteers, are captured via iRecord for your taxon group(s) would you be willing for it to be transferred directly to the NBN Gateway?

Yes	21%	(n=13)
No	13%	(n=8)
Maybe	14%	(n=9)

#### Comments

- Only if it is verified - otherwise reflects badly on time take to verify other records
- Need to clearly mark when verified with the measure of verification
- Depends on agreements at a local level (with county recorders etc)
- We would want all data to come to use (Recording Scheme) before going on NBN Gateway for verification - data is more useful to the Recording Scheme is it flows through our own database.
- Would like to be notified when data are added
- Would prefer records go via LERCs to the Gateway as that reduces the time before records because realistically available to local planners and reduces the admin burden of LERC if they had to trawl the Gateway/iRecord for data.

### Q10. Repatriating data

If you curate data on the NBN Gateway that is for non-UK areas (e.g. Republic of Ireland), would you be willing for that data to be repatriated to the country in which it was collected (e.g. a policy of Irish data held by the National Biodiversity Data Centre (for Ireland))?

Maybe	11%	(n=7)
No	11%	(n=7)
Yes	27%	(n=17)

#### Comments

- Would need more information
- Would need recorders/owners consent
- Already supply relevant data to NBDC and Isle of Man and Channel Islands record centres

### Q11. NBN Data Exchange with GBIF

Will you allow data to be shared with GBIF?

Maybe	35%	(n=22)
-------	-----	--------

No	8%	(n=5)
Yes	54%	(n=34)

If so, which is your preferred license?

CC0	8%	(n=5)
CC-BY	13%	(n=8)
CC-BY-NC	49%	(n=31)

### Comments

- Will need to check if CC0 is okay with government licensing
- Only if happy with arrangements. It may be that we only allow a subset of our data onto GBIF
- Need to review and check permissions for our third party data holdings/recorders/owners
- If the resolution and/or access of the data can be restricted.
- Need more resources to be able to do this
- Our preferred license option would be none of the licenses above as we would not be willing to allow access to all organisations via this license.
- I strongly want to share data for distribution purposes but not be forced into loss of access controls
- We need more time and information in order to understand the potential implications of these and to review the data and update our data policies
- Worried about the way GBIF displays data. All of our records are attributed to BNG, and relate to a whole grid square. GBIF displays all data as points. This will lead to records we hold being displayed at a false precision and in locations where the record was not made. This in turn could lead to misunderstandings and flawed interpretation/research.
- Would provide under CC0 provided that there are community norms for attribution as described during the GBIF consultation on licensing, unless there is any clear documentation demonstrating that there is a firm legal basis for monitoring and enforcing CC-BY. We are not happy that our data has stopped being transferred to GBIF and I see from comments on the NBN Forum that other data providers are not happy about this either. Some clarity over data flow to OBIS would be useful too.
- Our business plan does not currently take into account of the potential loss of income from carrying out data search requests. Further research would have to be undertaken to investigate the effects this may have.
- We would not be happy to share under those licences

## Q12. Access to OS Products

Would access to OS mapping products would be useful to you?

Yes	60%	(n=38)
-----	-----	--------

No 27% (n=17)

### Which products would be useful

- VectorMap Local
- Mastermap
- High resolution Rasters (VML, VMD)
- DMT and landform data such as OS terrain 5 and historic mapping
- All base maps (1:50000, 1:25000)
- Mastermap Topography and Imagery
- Datasets showing field boundaries
- All products
- All OS open data products
- OS Terrain 5
- Elevation and OS county series
- OS Explorer maps

**Comment** - Please don't exclude nautical charts from this project, please speak to UKHO as well as OS.

### What would you use them for?

- Precise mapping of habitats and sites
- Production of recording maps, displaying results, project design etc
- Routine LERC roles, webservices and projects
- Display of biological records for species and habitats on maps,
- Capturing accurate grid references for biological records,
- Digitising habitats
- Maps to support our conservation work and communicate the results from our survey and monitoring program
- Field data capture
- Verification of records
- Mastermap data would be useful for Phase 1 and habitat mapping, and for use with the ECOSERVE toolkit developed by the Wildlife Trusts, which we are unable to use currently due to lack of Mastermap data and staff resource to run the tool. Terrain data would be useful for habitat modelling. All of the products produced by these tools are currently unavailable to our partners because although we have the expertise, we cannot get the data to run them.
- Analysis - Network mapping, 3D modelling
- Deriving data (e.g. reserve boundaries)
- Improving the quality of recording and digitisation
- Basemaps in report writing to statutory agencies
- Use of terrain data in habitat management and habitat suitability analysis
- Keeping up to date database of freshwater ponds
- Online recording
- Atlas projects

### Q13. Vice-county Boundaries

**Are you happy with the accuracy of the Vice-County boundaries?**

Yes 67% (n=42)  
No 13% (n=8)

**If not, what errors or problems have you identified**

- Isles of Scilly in VC1 is a real problem - the start of VC144 made no sense and did not help.
- The missing chunk of Cornwall from VC2, due to the trib of the Tamar.
- Don't have any digitised boundaries for NI

**General comments**

- I would like to be informed if changes are made to the current published boundaries.
- Wow, you could really open a can of worms if you start revising vice-county boundaries
- I would certainly want confirmation of the validity of the boundaries before using and relying on them

**General any other comments**

- The instability issues with the Gateway should be a priority problem to address.
- The problem of data access continues to frustrate us. Records for which there is no public access or only available at 10km resolution is almost useless.
- We need to explore these ideas with our partners - local organisations, groups, individuals. Could some jargon-free information\* be produced centrally, for us to adapt to our local conditions, to stimulate debate, discussion and decision please? (\*e.g. powerpoint presentation or similar)
- Resources page on BRC website,
- Would be nice if you could just download VC boundaries or the full set + UK outline from there. Free GIS is so easy to use and the VC outlines + GR squares are the first thing that would help entice recording scheme organisers and other researchers to utilise it.
- Need a specific forum/web space for Biogeographical techniques & resources, the NBN forum is so cluttered.
- Need some means of submitting data outside UK (NBN don't reply, some countries don't have GBIF contributors - Hungary) how about BRC?

**Additional Comments:**

I'm still unclear how big organisations can quickly send and update data with the NBN. Can we fire in records into NBN via an API like iRecord can? I'd like our data to be passed across in an automated way, no email, no spreadsheet, just our database talking with your database and a load of validation rules spotting errors.

Please feel free to contact me if you wish to discuss any aspect of my answers. I would be particularly interested in any progress which could be made towards more openness generally although, if you accept that there will always be a need for some restrictions in some cases, which most people (although not everyone) does, then it would be useful to see if a more cohesive approach to what data is considered sensitive and how this is managed could be created.

The single biggest problem we have is lack of resources, particularly long term funding, and this is only likely to get worse as Government funding is reduced. This needs to be addressed if we are to be able to provide data to the NBN at all in future years. The other problem is that our non-Govt funders do not use the NBN, and so the Gateway needs to be promoted to them if we are to put data on there. Is the NBN currently investigating funding and accessibility of data for potential data users who do not currently access data via the Gateway? There is no point in us adding data to the Gateway if our main data users want it in a different format!

The NBN seem to be concentrating a lot on historic data at present, whereas our priority and our funders priority is recent data, therefore NBN work takes us away from our core source of income. If the NBN strategy was more aligned with that of our data users then we would use the Gateway more.

The other critical problem for us is that we use Indicia to collate data and speed up data flow, but the NBN and systems linking to it like Indicia do not receive enough support. There needs to be support for web systems and coding to speed up data flow, and the NBN should consider this as key to their strategy on improving access to data.

DGERC are fully supportive of making biological data more readily available to a variety of data users. The push for more open access to high-resolution data creates a dilemma, since this could potentially put at risk the business model on which that LERCs like DGERC are founded. Most (all?) LERCs rely to a greater or lesser degree on income from commercial organisations as well as from the public sector. Indeed with public funding diminishing, many LERCs are striving to diversify their funding streams. Future NBN developments (e.g. Atlas of Living ...) must not undermine the ability of contributing organisations to make ends meet. If LERCs are deemed a desirable and valuable asset to the network (and DGERC believes that their work in promoting/co-ordinating wildlife recording at a local level

is an essential part of delivering the NBN's future strategy), a mechanism supporting a sustainable business model for such Centres must be ensured before there can be more free and open access to data.

SxBRC, as a Local Environmental Record Centre, plays a key role in maintaining the network of County Recorders and providing support and training locally with iRecord verification. This represents a significant investment, by SxBRC, in ensuring data is quality assured at the county level. We welcome the dialogue NBN is having with the national recording schemes, to look at the data flow from iRecord to the NBN. It would be great to be kept informed about how this goes. Please can the value that LERCs add to this process be considered and accounted for in coming to any decisions. Thanks!

I am personally very much pro-OpenData and Merseyside BioBank will continue to more and support that as a direction. We support opendata as a principle and absolutely understand the benefits. There are two points however that unless tackled will prevent us from achieving this goal;

Commercial use: MBB draws a small income from commercial use and services which covers around 15% of our running costs. A very small amount for most organisations. However, that income demonstrates that there is a continued need and use of the LRC and allows us to cover the rest of our running costs through funding from the local authorities we work with. In turn that allows us to continue to facilitate and support existing and new recorders and recording groups. Training volunteers skills that assist them into employment and develop their interests and specialisms, supporting local groups to enhance their membership and draw down funding for their projects. In losing commercial income we would cease to receive support from the local planning authorities and no longer be able to provide these services and much of the data and added benefits of the LRC (many of which deliver on NBN strategic goals) would be lost.

Provider consent: MBB works on the basis that the records we hold remain the property of the original provider. While we invest significant time and resources in the digitisation and management of these data it is a matter of trust that organisations continue to work with us. Some of the organisations we work with would not be willing for the records they share to be made fully open. As a case in point the local ARG group has recently been very concerned when ARC made their rare species dataset available at 100m resolution. This included a large amount of Sand Lizard data, a species which is highly protected and prone to collection. The ARG were not consulted before hand may now stop sharing information with ARC.

Another gap which NBN could help with to facilitate more records being shared on the NBN-G is in the technical capacity of LERC staff to properly manage online

recording websites. Many LERCs are using Indicia-based online recording websites, but do not have the staff capacity to maintain all the updates, create new pages for recording groups, develop new functionality etc. Ideally the technical skills would be kept in-house and existing LERC staff would be trained to the required level, or new staff would be employed who already have the technical skills required. However, this will take time and in the short-term there's also a gap in the number of private consultants who have the skills to help LERCs with this while existing LERC staff gain the considerable technical skill required.

More generally, the comments in this questionnaire are from my perspective as organiser of the national recording scheme for soldierflies and allies. I am also a county recorder for moths in Berkshire. Personally, I would like to see the options I have ticked above apply to my county moth data as well, although my understanding is that Butterfly Conservation, as custodians of the national moth recording scheme, do not take the same view as me. I would hope that the moth data can also become much more openly available in future, without compromising BC's ability to fund and run the recording scheme.

If a UK Atlas becomes the accepted way for all potential users to access data for all kinds of uses, then recognition and sustainable financial support needs to be provided to the many organisations such as LRCs and recording schemes that spend a lot of time collating, checking, validating and managing data for mobilisation to the NBN Gateway. Researchers, academic bodies and commercial operations should not expect to use data for free unless the data providers are adequately and sustainably funded by government or a common fund that is paid into by Researchers, academic bodies and commercial operations researchers, academic bodies and commercial operations.

ARC shares a vision of open biodiversity data, and we feel that the answers to most of these survey questions is a qualified yes. However, as an NGO we are concerned that the current landscape sees some data providers rather sidelined by the statutory agencies in terms of support for data management, survey co-ordination and volunteer engagement. It's a competitive field, with important organisational issues and the NBN Trust needs to recognise this. We very much support the move towards an Atlas of Living Australia infrastructure and a mechanism for improved access of Ordnance Survey data to the charity sector.

This whole plan seems flawed and designed to undermine the goodwill of our recorders and LRCs. We do not make our all our data open for good reason; because some species are sensitive, some recorders do not want their data freely available and some data is commercially sensitive.